

**APPLICATION  
FOR  
UNITED STATES LETTERS PATENT**

APPLICANT NAME: Prakash

TITLE: EDGE SMOOTHING FILTER FOR CHARACTER  
RECOGNITION

DOCKET NO.: CHA920030036US1

**INTERNATIONAL BUSINESS MACHINES CORPORATION**

**CERTIFICATE OF MAILING UNDER 37 CFR 1.10**

I hereby certify that, on the date shown below, this correspondence is  
being deposited with the United States Postal Service in an envelope  
addressed to Mail Stop Patent Application, Commissioner for  
Patents, P.O. Box 1450, Alexandria, VA 22313-1450 as  
"Express Mail Post Office to Addressee"  
Mailing Label No. EV 393 299 375 US

on February 17, 2004

Beverly Kehoe Shea

Name of person mailing paper

Beverly Kehoe Shea  
Signature

February 17, 2004

Date

# EDGE SMOOTHING FILTER FOR CHARACTER RECOGNITION

## BACKGROUND OF THE INVENTION

### 1. Technical Field

The present invention relates generally to character recognition systems, and more specifically relates to an edge smoothing filter for correcting defects in black white image details.

### 2. Related Art

The ability to read and interpret character information remains an important challenge for enterprises, such as banks, that must process printed documents. In a typical application, documents such as bank checks are scanned and then character information, such as account numbers, etc., is extracted and stored. Often, during the scanning process, defects are introduced that lead to errors in the character recognition process.

One cause for such defects results from the bitonal conversion algorithms used to generate black white images. In most scanners, black white images are generated from gray scale images. A bitonal conversion algorithm is used to convert the gray scale image to a black white image, which is then read by a character recognition system, such as OCR (Optical Character Recognition), MICR (Magnetic Ink Character Recognition), thermal imaging, etc.

Figure 1 depicts a black white image with character data (i.e., string “6 6”) containing two types of defects that are commonly introduced. As can be seen, various edges of the characters have protrusions 12 (i.e., black pixels protruding out) and/or indentations 14 (i.e., black pixels missing). More often than not, several such defects are present in character data found in black white images. These types of defects can create problems for character recognition, image scaling, etc. Accordingly, a need exists for a system and method for eliminating character defects in black white images.

## **SUMMARY OF THE INVENTION**

The present invention addresses the above-mentioned problems, as well as others, by providing an edge smoothing filter for correcting defects in character data of a black white image. In a first aspect, the invention provides an edge smoothing filter for correcting defects in a black white image, comprising: a system for processing blocks of pixels in the black white image, wherein each block comprises one center pixel and eight exterior pixels; and an algorithm that examines each block of pixels, wherein the algorithm overwrites the value of the center pixel if: all three pixels along a first edge share a first value; and all three pixels along an opposing edge share a second value that is opposite of the first value; and the two exterior pixels residing between the first and second edge share a common value.

In a second aspect, the invention provides a method for correcting defects in a black white image, comprising: selecting a block of pixels from the black white image, wherein the block comprises one center pixel and eight exterior pixels; examining the block of pixels to determine if: all three pixels along a first edge share a first value, and

all three pixels along an opposing edge share a second value that is opposite of the first value, and two exterior pixels residing between the first and second edge share a common value; and if the above conditions are met, overwriting the value of the center pixel with the common value of the two exterior pixels.

In a third aspect, the invention provides a program product stored on a recordable medium for correcting defects in a black white image, comprising: means for selecting a block of pixels, wherein the block comprises one center pixel and eight exterior pixels; and means for examining the block of pixels to determine if: all three pixels along a first edge share a first value, and all three pixels along an opposing edge share a second value that is opposite of the first value, and two exterior pixels residing between the first and second edge share a common value; and means for overwriting the center pixel with the common value of the two exterior pixels if all conditions of the examining means are met.

In a fourth aspect, the invention provides a character recognition system that processes character data from a black white image, comprising: an edge smoothing filter that examines blocks of pixels in the black white image to determine if an indent or protrusion defect exists, and if so corrects the defect in a filtered image; and a recognition engine examines the filtered image and extracts character information.

#### **BRIEF DESCRIPTION OF THE DRAWINGS**

These and other features of this invention will be more readily understood from the following detailed description of the various aspects of the invention taken in conjunction with the accompanying drawings in which:

Figure 1 depicts black white character data having defects.

Figure 2 depicts a character recognition system having an edge smoothing filter in accordance with the present invention.

Figure 3 depicts the black white character data of Figure 1 after it has been passed through the edge smoothing filter of Figure 2.

## **DETAILED DESCRIPTION OF THE INVENTION**

Referring now to Figure 2, an exemplary implementation of the edge smoothing filter 30 of the present invention is described embodied in a character recognition system

10. In this embodiment, a character recognition system 10 is provided that receives a black white image 28 containing printed characters or character data. Before being processed by a recognition engine 34, the black white image 28 is first passed through the edge smoothing filter 30 to generate filtered black white image 32. The filtered black white image 32 is then processed by one or more recognition engines 34 to collect character information 35, which can then be further processed, stored in memory 36, etc.

In this exemplary embodiment, a document 20 is scanned by a scanner 22, which first generates a gray scale image 24. The gray scale image 24 is then processed using a bitonal conversion algorithm 26 to generate black white image 28. It is recognized that systems for generating black white images in this manner are known in the art, and therefore such systems are not discussed in further detail. It should also be understood that black white image 28 can originate from sources other than scanner 22, e.g., it could be retrieved from memory, etc.

As noted above, character data in black white image 28 often contains protrusion and indentation defects that can lead to errors by the recognition engine 34. Edge smoothing filter 30 seeks to correct these types of defects by examining blocks of pixels in the black white image 28, and if necessary, it overwrites pixel values in accordance with a set of rules embodied in algorithm 31. Edge smoothing filter 30 works by examining blocks of nine pixels arranged as follows, wherein each pixel has a value of either black or white:

p1	p2	p3
p4	p5	p6
p7	p8	p9

Depending on the values of the exterior pixels (p1-p4 and p6-p9) the algorithm 31 may automatically overwrite a value of the center pixel p5. Filter 30 sweeps through the entire black white image 28 in a rasterized manner such all of the pixels in the image, except for those pixels along the outer edges of the image, are selected as the center pixel p5 in the block of pixels. Note that any method may be used for sweeping through the black white image 28, e.g., a horizontal sweep, a vertical sweep, etc. Figure 3 depicts the character data from Figure 1 after being processed by edge smoothing filter 30. As can be seen, indentation and protrusion defects have been eliminated.

Each time a new block is selected by filter 30 during the sweep, algorithm 31 is applied to determine whether the center pixel p5 should be overwritten. Algorithm 31 can be summarized as follows:

Select a block of nine pixels having one center pixel and eight exterior pixels, and perform the following test:

If (all three pixels along a first edge share a first value) and (all three pixels along an opposing edge share a second value that is opposite of the first value) and (the two exterior pixels between the first and second edge share a common value), then rewrite the center pixel to the common value of the two exterior pixels.

A four-step procedure for implementing algorithm 31 is explained below with reference to Tables 1 and 2. Table 1 depicts a 3x3 cell, with the center pixel P being the pixel of interest. Each entry in the 3x3 cell is a pixel location.

Table 1

L	T	R
L	P	R
L	B	R

Depending on the characteristics of the eight neighboring pixels, the pixel P may be overwritten with a black or white pixel.

Step 1. Are the three left cells (L) of the same value (i.e., either all white or all black) and are all three right cells (R) of the same value, but opposite to the left cells (e.g., all L's are black and all R's are white)? If yes, go to step 2, else go to step 3.

Step 2. Are both exterior pixels between the left and right edges (i.e., the top pixel T and the bottom pixel B) of the same value? If yes, overwrite P with that value and exit. If no, go to step 3.

Step 3. Reconfigure the 3x3 cell to that of Table 2.

Table 2

T	T	T
L	P	R
B	B	B

Are the top three cells T of the same value and are the bottom three cells B of the same value but opposite the three top cells T? If yes, go to step 4, else exit.

Step 4. Are both exterior pixels between the top and bottom edges (i.e., left pixel L and the right pixel R) of the same value? If yes, overwrite P with that value, else exit.

Consider the following defective pixel layout examples where B is a black pixel and W is a white pixel.

**Example 1:**

Consider the pixel layout that includes a protrusion:

B	W	W
B	B	W
B	W	W

Using the four-step process described above, the answer to step 1 would be yes since the left three pixels are of the same value (i.e., black) and the right three pixels are of the same value (i.e., white) but opposite the left three. The answer to step 2 would also be yes since the top and bottom (middle) pixels are both white. Thus, the middle pixel would be overwritten with a value white, as follows:

B	W	W
B	W	W
B	W	W

As can be seen, the protrusion in the original pixel layout has been smoothed out.

**Example 2:**

Consider the following pixel layout having an indentation:

B	B	B
B	W	B
W	W	W

In this case, the answer to step 1 would be no, which would send the algorithm to step 3. The answer to step 3 would be yes since the top three pixels are of the same value (i.e., black) and the bottom three pixels are of the same value (i.e., white) but opposite the top three. The answer to step 4 would also be yes since the left and right (middle) pixels are both black. Thus, the middle pixel would be overwritten with a value black, as follows:

B	B	B
B	B	B

W W W

As can be seen, the indentation in the original pixel layout has been smoothed out.

It should be understood that black white image 28 may comprise a limited portion of a larger image, e.g., it may comprise the part of a bank check that just includes text data such as the account number, etc. Furthermore, it is understood that edge smoothing filter 30 need not be integrated with character recognition system 10, i.e., it could be implemented as a separate standalone system, be integrated into scanner 22, be utilized with a scaling system, etc.

It is understood that the systems, functions, mechanisms, methods, and modules described herein can be implemented in hardware, software, or a combination of hardware and software. They may be implemented by any type of computer system or other apparatus adapted for carrying out the methods described herein. A typical combination of hardware and software could be a general-purpose computer system with a computer program that, when loaded and executed, controls the computer system such that it carries out the methods described herein. Alternatively, a specific use computer, containing specialized hardware for carrying out one or more of the functional tasks of the invention could be utilized. The present invention can also be embedded in a computer program product, which comprises all the features enabling the implementation of the methods and functions described herein, and which - when loaded in a computer system - is able to carry out these methods and functions. Computer program, software program, program, program product, or software, in the present context mean any expression, in any language, code or notation, of a set of instructions intended to cause a system having an information processing capability to perform a particular function

either directly or after either or both of the following: (a) conversion to another language, code or notation; and/or (b) reproduction in a different material form.

The foregoing description of the preferred embodiments of the invention has been presented for purposes of illustration and description. They are not intended to be exhaustive or to limit the invention to the precise form disclosed, and obviously many modifications and variations are possible in light of the above teachings. Such modifications and variations that are apparent to a person skilled in the art are intended to be included within the scope of this invention as defined by the accompanying claims.